

UTILIZACIÓN DE MOODLE COMO PLATAFORMA PARA LA INVESTIGACIÓN EDUCATIVA: APLICACIÓN A LOS CÓRPORA DE APRENDICES DE LENGUAS.

USING MOODLE AS A PLATFORM FOR EDUCATIONAL RESEARCH: AN APPLICATION TO COMPUTER LEARNER CORPORA.

Dr. Antonio Moreno Ortiz
amo@uma.es

Dra. Chantal Pérez Hernández
mph@uma.es

Elena del Olmo Bañuelos
eob@uma.es

*Universidad de Málaga. Facultad de Filosofía y Letras.
Departamento de Filología Inglesa, Francesa y Alemana.
Campus de Teatinos s/n, 29071, Málaga (España)*

En este artículo describimos la infraestructura y protocolos que hemos desarrollado para utilizar el gestor de aprendizaje Moodle como plataforma de apoyo a la investigación, en este caso para la creación de un corpus de aprendices de lenguas extranjeras, cuya finalidad es el estudio del conocimiento que los estudiantes universitarios tienen de la lengua inglesa. Ponemos el énfasis en la metodología empleada, más que en la aplicación específica a los corpóra de aprendices de lenguas, ya que pensamos que puede ser aplicada a cualquier otro ámbito de investigación educativa que requiera de un proceso de recogida de datos controlado.

Palabras clave: Recogida de datos, sistemas de gestión del aprendizaje, inglés como lengua extranjera, corpóra informatizados de aprendices, minería de datos educativa.

This paper describes the infrastructure and protocols that we have developed in order to use the Moodle Learning Management System as a platform for research support, specifically for the creation of a Computer Learner Corpus, whose aim is to study the English language skills of university students. The emphasis is on the methodology itself, rather than its specific application to Learner Corpora studies, since we believe that it can be applied to any other topic in educational research that requires a controlled data collection process.

Keywords: Data collection, learning management systems, English as a foreign language, computer learner corpora, educational data mining.

1. Introducción.

Una de las partes más tediosas y laboriosas en el proceso de investigación en diversos ámbitos de conocimiento es la fase de recogida de datos. Algunos tipos de instrumentos, como el cuestionario, han encontrado en las TIC diversas formas de facilitar su implementación y ejecución, cambiando el medio tradicional impreso por el electrónico y desarrollando el proceso de recogida de datos a través de Internet. Este cambio en el modo de aplicar la metodología sin duda alguna tiene sus ventajas, aunque también sus inconvenientes, entre los que destacan los relativos a cobertura de la población, selección de muestra y tasas de respuesta (Vicente & Reis, 2007; Estévez, 2012).

En lo que respecta a la enseñanza de lenguas extranjeras, la lingüística de corpus ha desarrollado metodologías específicas para el análisis de las producciones lingüísticas de los estudiantes de una lengua mediante los denominados *Córpore informatizados de aprendices* (*Computer Learner Corpora*), que son recopilaciones de textos y otros materiales producidos por los estudiantes, almacenados en formato digital y anotados, según determinados esquemas, para facilitar la recuperación y el análisis de los datos textuales (Granger, 2002). No obstante, la compilación de estas valiosas fuentes de datos ha supuesto tradicionalmente un proceso complejo y difícil, ya que, aunque en principio implica los mismos pasos que la aplicación de cualquier instrumento de investigación, convertir los datos de formato impreso a electrónico supone una cantidad de trabajo adicional muy considerable por parte del investigador (Cestero, Penados, Blanco, Camargo & Granda, 2001; Granger, 2002; Torruella & Llisteri, 1999), pues conlleva la

transcripción textual, palabra por palabra, de las producciones escritas de los estudiantes. Por ello, el empleo de una plataforma on-line en la que los estudiantes puedan generar los textos que conformarían dicho corpus de estudio es una alternativa especialmente atractiva para el investigador de este campo, pues simplificaría enormemente este proceso.

Por otra parte, el aprovechamiento de los gestores de aprendizaje on-line como recurso del que obtener datos para la investigación educativa no es nuevo (Dougiamas, 2003; Ros Martínez de Lahidalga, 2008). De hecho, en los últimos años se ha consolidado un campo de estudio denominado *Minería de datos educativa*, que es una especialización de la minería de datos que se centra en la explotación de los contenidos generados en entornos de aprendizaje on-line, con el objeto de descubrir información relevante (Romero & Ventura, 2010). En palabras de Baker y Yacef (2009, p. 4) «Educational data mining (also referred to as EDM) is defined as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in». Por ejemplo, Romero, Ventura y García (2008) describen un procedimiento para extraer información de Moodle empleando técnicas de minería de datos y minería de textos, aplicando también técnicas de aprendizaje automático. Este campo, que cuenta ya con su propia revista especializada, sin duda tiene mucho que ofrecer a la comunidad educativa, ya que el uso de las plataformas colaborativas en la Educación es algo generalizado (Dobozý & Reynolds, 2010). De hecho, la generalización de estos entornos, y por extensión el resto de la Red, queda patente en la naturaleza misma

de las últimas tendencias en teorías de aprendizaje, como es el conectivismo (Siemens, 2005), por lo que sería un gran desperdicio no aprovechar estos recursos tecnológicos y los contenidos que en ellos se generan.

Nuestro enfoque no se basa en la *Minería de datos educativa*, sino en la creación de contenidos mediante actividades dirigidas y planificadas al efecto, pues el tipo de datos que buscamos recopilar no se produce de manera *natural* en las actividades realizadas on-line, ya que los alumnos, por lo general, no generan textos en lengua extranjera en el transcurso normal de sus estudios universitarios. Además, hemos pretendido crear un entorno controlado en el que se minimicen los riesgos mencionados en cuanto al uso de cuestionarios y otros instrumentos on-line, que permita obtener datos válidos de una manera eficiente y sistemática.

El proceso y metodologías que aquí describimos lo hemos desarrollado dentro del proyecto I+D+i FFI2011-25893, denominado «Lingmotif: Análisis de Sentimiento en Textos: Aplicaciones Didácticas y Financieras». Una de las tareas a desarrollar en este proyecto consiste en la compilación de textos de discurso evaluativo para el inglés; por tanto, el corpus de aprendices que presentamos en este artículo aprovecha esta infraestructura de recursos informáticos, lingüísticos, humanos y de conocimiento desarrollados en dicho proyecto. La recopilación de textos escritos en inglés para las tareas de análisis de sentimiento, que conforman el objetivo general del proyecto, nos servirá como corpus de referencia de lengua inglesa nativa para su comparación con el corpus de aprendices de inglés como lengua extranjera; y es este tipo de estudio contrastivo en el uso de un idioma como lengua materna y como lengua extranjera

el que constituye la base de «Contrastive Interlanguage Analysis» (Granger, 2004, p. 132).

Como complemento importante al corpus, nos planteamos, además, la recopilación de datos que ayuden a extraer mayor información acerca del conocimiento de la lengua inglesa de los alumnos universitarios. Para ello, hemos diseñado otros dos instrumentos de recogida de datos: un cuestionario, cuyo objetivo es obtener datos objetivos relativos a la experiencia en el aprendizaje del inglés (número de años, posibles influencias externas en su conocimiento de la lengua, etc.), y subjetivos, relativo a las percepciones que los aprendices tienen sobre su proceso de aprendizaje de lenguas extranjeras; y un test estandarizado, elaborado por una reputada institución británica, comúnmente empleado para la asignación de nivel de conocimiento del inglés. En la apartado 4 describimos los tres instrumentos de recogida de datos.

Como hemos apuntado, nuestra propuesta para la implementación y ejecución de estos tres instrumentos (cuestionario, test y ensayo) es aprovechar las funcionalidades de la plataforma Moodle y emplearla como herramienta para la investigación (educativa). En este sentido, es preciso advertir que el uso que hacemos de Moodle en el trabajo que aquí presentamos supone un cambio conceptual en lo que se refiere a los elementos que componen un gestor de aprendizaje (LMS: Learning Management System) (Correa, 2005; Cobb & Steele, 2011; Mallon, 2011): nuestros cursos son simplemente contenedores de instrumentos de recogida de datos; los profesores son los investigadores que diseñan los instrumentos y los implementan dentro del curso; y los alumnos son los participantes en la investigación, que en el caso que aquí presentamos son, de hecho, alumnos

universitarios.

En el resto del trabajo detallamos el proceso seguido para llevar a cabo estos objetivos, que sintetizamos aquí como guía de referencia:

1. Instalación del servidor y el software.
2. Puesta en marcha del LMS y prueba de funcionamiento.
3. Diseño de los instrumentos y su implementación como actividades en el LMS.
4. Proceso de recogida de datos.
 - a) Proceso de registro de estudiantes.
 - b) Obtención del input de los estudiantes.
5. Exportación de los datos a formatos compatibles con paquetes de software para análisis.

La implementación de Moodle que describimos se encuentra accesible en la dirección <http://www.tecnolengua.uma.es>. Es posible darse de alta en la misma y solicitar más información sobre la posibilidad de usarla para fines académicos por parte de investigadores de las universidades españolas.

2. Córpora de aprendices de lenguas extranjeras.

Existe una extensa bibliografía que da cuenta de las propuestas teóricas y aplicadas y los progresos hechos desde los años 60 en el ámbito del análisis de errores -véase, por ejemplo, Ellis (1997), Lengo (1995) o Brown (2000)- y, como es natural, la aparición de las nuevas tecnologías ha facilitado en cierta medida el proceso de recogida de datos, identificación y clasificación de errores (Dagneux, Dennes, & Granger, 1998; Granger, 2003a). Sin embargo, sigue siendo un proceso costoso en términos de tiempo y recursos humanos y, como señala Tono (2003), la asignación de etiquetas a los errores cometidos por los alumnos para su

clasificación y estudio requiere un marco teórico que sistematice el trabajo y, por supuesto, herramientas informáticas que permitan recoger los datos de forma controlada y analizarlos de forma versátil.

La creación de córpora de aprendices de lenguas extranjeras nació como unión de los presupuestos teóricos y metodológicos de la *Lingüística de Corpus* y de los estudios de *Adquisición y Enseñanza de Lenguas Extranjeras* (Granger, 2002). La *Lingüística de Corpus* basa sus estudios (ya sean gramaticales, léxicos, semánticos, pragmáticos o de cualquier otra índole) en el análisis de muestras reales de uso lingüístico, recogidas en un corpus textual informatizado, que puede procesarse de diferentes maneras para obtener evidencias sobre la forma en la que los hablantes de una lengua crean y estructuran sus producciones lingüísticas (Pérez Hernández, 2002). Dichos córpora se crean siguiendo una serie de criterios lingüísticos y extralingüísticos (Atkins, Clear, & Ostler, 1992; Tono, 2003), y se les puede añadir información morfológica, sintáctica o de otros tipos, que ayuda en el proceso de análisis y en la identificación exacta de las circunstancias en las que fueron producidos (tipo de texto, fecha, lugar, tipo de hablante, etc.) y compilados (fecha de inclusión en el corpus, proceso de limpieza/etiquetado de datos, etc.) (O'Keeffe & McCarthy, 2010). Un corpus, por tanto, se construye como muestra representativa del uso que los hablantes hacen de una lengua o de una variedad lingüística determinada y su estudio ha revelado, por ejemplo, que existen patrones regulares de uso y asociaciones entre palabras que no son obvios ni siquiera para los hablantes nativos de una lengua si usan su intuición lingüística como única herramienta de análisis (Sinclair 1991, 2004).

En cuanto a los corpóra creados con textos producidos por aprendices de una lengua, hace ya más de una década que Tono (2003) destacó su utilidad para la investigación sobre la adquisición y enseñanza de lenguas extranjeras. A pesar de que los corpóra de aprendices hayan sido considerados como un tipo de corpus especializado (Römer, 2008), sigue siendo un reflejo del uso real del lenguaje en un contexto educativo; además, como reconoce Granger (2004) su estudio está aportando nuevas ideas y rebatiendo concepciones previas que se tenían sobre el proceso de aprendizaje de una lengua extranjera ya que permite, por ejemplo: describir las distintas etapas del desarrollo de *interlanguage*, en español *interlengua*, un término acuñado por L. Selinker (1972) y que hace referencia al sistema lingüístico que posee un estudiante de una segunda lengua o lengua extranjera en cada una de las etapas del proceso de aprendizaje; identificar errores ocasionados por la transferencia de estructuras o palabras de la lengua materna del aprendiz (Roca, Piñeiro & Gigirey, 2012); evidenciar el uso excesivo, o escaso, de determinados patrones lingüísticos de la *interlengua*; distinguir entre errores universales, comunes a todos los aprendices de un idioma, y errores determinados por la lengua materna de los aprendices; comparar patrones lingüísticos regulares entre el uso que hacen los hablantes nativos y los aprendices de la lengua objeto.

3. Implementación de Moodle como plataforma de apoyo a la investigación.

El primer planteamiento que hemos de hacernos es si existe la necesidad de implementar nuestro propio sitio Moodle o bien podemos emplear una implementación

existente. En nuestro caso, la Universidad de Málaga dispone de una implementación a gran escala que da cabida a todas las necesidades de formación on-line de la institución a nivel de Grado, Posgrado, Formación del Personal Docente e Investigador, Personal de Administración y Servicios y otros. Tener acceso a una implementación consolidada como ésta tiene sus obvias ventajas, pero también importantes inconvenientes.

Las primeras se refieren a que no es necesario tener grandes conocimientos técnicos, sino simplemente ser usuario avanzado de Moodle en cuanto a la creación de contenidos y exportación de datos. Por otra parte, los inconvenientes son muchos: en primer lugar, hemos de tener privilegios de administrador para poder llevar a cabo un buen número de tareas necesarias, o bien contar con apoyo de personal de soporte técnico dispuesto a dedicar una cantidad tiempo considerable a nuestros intereses. Algo que puede suponer un obstáculo insalvable para optar por esta posibilidad es el sistema de control de acceso de usuarios de la institución. En nuestro caso, el campus virtual de la Universidad de Málaga permite el acceso únicamente a usuarios registrados en la Universidad de Málaga (alumnos, PDI o PAS), algo lógico por motivos de seguridad, pero que impediría la recogida de datos de alumnos de fuera de la propia institución. Por tanto, esta idea quedó descartada desde el principio, ya que pretendemos llevar a cabo estudios de mayor ámbito geográfico. Finalmente, siempre estaríamos limitados en cuanto a las posibilidades de personalización del sitio. La otra alternativa consiste en desplegar nuestra propia implementación de Moodle, en cuyo caso no tendríamos ninguna de las limitaciones que mencionamos, pero, obviamente, requiere de determinados

recursos y conocimientos técnicos. En lo que resta de esta sección, nos proponemos aportar una idea de lo que implica llevar a cabo esta iniciativa.

3.1. Requisitos de hardware y software.

Las necesidades en cuanto a hardware dependen de una serie de factores, entre los que destacan el número de usuarios potenciales, especialmente de forma concurrente y el ancho de banda disponible, que garantice un tiempo de latencia aceptable para evitar cuellos de botella que podrían resultar en una pérdida de datos. Nuestra implementación se ejecuta sobre un servidor de arquitectura Intel Xeon con ocho núcleos y almacenamiento redundante RAID, alojado en la Sala de Ordenadores del Servicio Central de Informática de la Universidad de Málaga. La conexión a RedUMA, que a su vez forma parte de la Red Informática Científica de Andalucía (RICA), garantiza un ancho de banda óptimo. Aunque aún no lo hemos comprobado, suponemos que el resultado sería el mismo utilizándolo fuera de la Comunidad Autónoma, ya que RICA forma parte de la Red académica y científica nacional (RedIRIS).

En cuanto al software, Moodle está pensado para desplegarse en una plataforma LAMP, es decir, GNU/Linux como sistema operativo, Apache como servidor web, MySQL como sistema gestor de bases de datos y PHP como entorno de programación. LAMP es una plataforma empleada en multitud de soluciones informáticas centradas en el código abierto, de probada eficacia y plenamente compatible con la filosofía Moodle. En nuestro caso, ésta es la opción elegida, aunque también es posible su despliegue en una plataforma WAMP, es decir,

con Microsoft Windows como sistema operativo. En nuestra experiencia, Linux es la opción óptima por varios motivos, en primer lugar por su probada fiabilidad, siendo utilizado por miles de servidores en todo el mundo; en segundo lugar, porque la mayoría de las implementaciones de Moodle existentes utilizan esta misma plataforma y, aunque mucha de la documentación existente sea aplicable a otros sistemas operativos, el grueso de la misma se refiere a Linux, por lo que siempre es más fácil obtener respuestas para este sistema. Finalmente, Linux, en cualquiera de sus muchas distribuciones disponibles (Ubuntu Server en nuestro caso), es particularmente apropiado para la gestión remota del servidor. Por otra parte, no hay que olvidar que, al igual Moodle, todos los componentes de la plataforma LAMP están disponibles bajo licencia GPL, por lo que el coste económico de todo el software necesario es cero.

3.2. Instalación y configuración.

En general, se puede decir que poner en marcha un sitio Moodle requiere de ciertos conocimientos técnicos más o menos avanzados. En cuanto al soporte, Moodle cuenta con una larga historia como proyecto de código abierto, contando con una amplia documentación on-line, que además se complementa con el contenido generado por sus muchos miles de usuarios mediante los foros alojados en el sitio web de Moodle. El proceso de instalación está descrito en Moodle.org, por lo que no la repetiremos aquí. Sí es importante tener en cuenta que la fase más compleja del proceso no es la instalación en sí misma, sino la configuración de los diversos servicios necesarios para un correcto funcionamiento: el servicio Cron, responsable

de muchas tareas de mantenimiento automatizadas, las copias de seguridad y, sobre todo, el sistema de notificación por e-mail, que requiere la instalación y configuración de Postfix, a no ser que se disponga de un servidor de correo externo previamente.

El siguiente paso consiste en la configuración de las categorías y cursos que se van a emplear para la recogida de datos. Esta configuración no es otra que el modo en que Moodle organiza sus contenidos: la unidad básica es el curso, al cual se asignan una serie de usuarios con uno o más roles (profesor, alumno, etc.); a su vez, los cursos se organizan en categorías y, opcionalmente, subcategorías. Una vez existe el curso y se asignan el o los usuarios con privilegios de edición (es decir, se les asigna el rol de profesor), se puede proceder al diseño de los instrumentos que se emplearán para la recogida de datos, y que describimos en el apartado 4, que a los ojos de Moodle no son sino actividades.

3.3. Seguridad y registro de usuarios.

Un aspecto crucial de todo sistema on-line es la seguridad, especialmente en el caso de sistemas como el que describimos en este trabajo, que no sólo permiten, sino que exigen el registro de usuarios. Moodle ofrece diversas estrategias para el registro de usuarios, desde permitir el acceso a invitados hasta el registro manual por parte del administrador del sistema, pasando por el auto registro mediante confirmación por e-mail. Estas opciones se pueden habilitar o deshabilitar temporalmente y también reforzar de distintas maneras. Por otro lado, hay que distinguir entre el registro en el LMS, a nivel de sistema y el registro (o matriculación) en los cursos que dan acceso

a los contenidos (los instrumentos de recogida de datos, en nuestro caso).

El modo que hemos implementado el sistema de acceso se basa en la aplicación del sistema de auto registro con confirmación mediante e-mail, en cuanto al alta del usuario en la plataforma, y la matriculación en el curso correspondiente mediante contraseña. Una característica muy interesante que aporta la última versión de Moodle es la asignación de usuarios dentro de un curso a un grupo de manera selectiva según la palabra clave empleada. Esto nos permite crear grupos independientes e invisibles entre sí dentro de un mismo curso, lo que facilita una gestión automatizada de los distintos grupos de alumnos que participan en la creación de nuestro corpus, quedando identificados mediante una etiqueta que asignamos a cada grupo, y al mismo tiempo se mantiene el conjunto de todos los alumnos participantes. Existe un tercer nivel de agrupación de usuarios superior al curso que aporta la versión 2.3 de Moodle, que es la cohorte y que se aplica a nivel global de la plataforma. Hemos empleado este tercer nivel para distinguir grandes grupos dentro del LMS, ya que lo empleamos también para tareas más tradicionales de apoyo a la docencia en diversos contextos. Todos los alumnos registrados para participar en la creación de nuestro CLC pertenecen a la cohorte CLC-Users, mientras que utilizamos otras convenciones para crear las etiquetas que nombran los grupos, que reflejan la Universidad de la que proceden, la Facultad o Centro, la titulación en la que estudian, el curso, etc. Esto nos sirve para poder gestionar los diversos grupos desde el propio Moodle, antes de proceder a la exportación de datos, aunque después, una vez exportados también contemos con estos campos de información

proporcionados por los propios usuarios. Como medida de seguridad adicional, hemos habilitado el sistema reCAPTCHA.

Aunque pueda parecer un proceso largo, en realidad se completa en escasos dos minutos. El envío del e-mail de confirmación es inmediato y, por otra parte, es un proceso al que el usuario habitual de Internet está más que acostumbrado.

4. Diseño de los instrumentos en Moodle.

En este apartado describimos los tres instrumentos de recogida de datos que hemos diseñado para este trabajo de investigación sobre CLC. Cabe mencionar que antes de la realización de las tres partes de la prueba (cuestionario, test y ensayo) todos los participantes firmaron un *consentimiento informado*, también disponible en la plataforma virtual, en el que se describe la naturaleza del estudio y de su propia participación.

4.1. Cuestionario.

El cuestionario consta de un total de 14 ítems y fue diseñado para obtener datos objetivos de tipo personal (p. e.j., nombre, edad, sexo), académico (p. e., facultad, departamento, experiencia docente), lingüístico (lengua materna, conocimiento de lenguas extranjeras, miembros familiares bilingües en inglés) y, más concretamente, datos referidos a la experiencia en el aprendizaje del inglés (número de años dedicados al aprendizaje de inglés, experiencia en algún modelo de enseñanza en el que el inglés sea la lengua vehicular, etc.); y, por otra parte, datos subjetivos que reflejan las percepciones sobre su propio nivel de inglés (autoevaluación de destrezas), la instrumentalidad del inglés, la enseñanza de

contenidos a través del inglés, etc. Con el fin de asegurar la validez de los datos obtenidos, ofrecemos una versión del cuestionario en inglés para aquellos alumnos extranjeros que no tengan un nivel suficiente de español.

En cuanto a la creación del instrumento en la plataforma, la instalación básica de Moodle ofrece los módulos de actividades más usuales, pero no incluye el cuestionario como tal, aunque parezca lo contrario si se usa la interfaz en español. Se trata de un problema terminológico: el término empleado en el original en inglés para la actividad instalada en el paquete básico es *quiz*, que es un módulo diseñado para implementar pruebas o exámenes. La traducción al español que se hizo fue la de *cuestionario*, quizás por evitar las connotaciones negativas, y en cierto modo contrarias a la filosofía del constructivismo social, de la palabra examen. La voz inglesa para el instrumento de encuesta por cuestionario es *questionnaire*, que es el nombre del módulo opcional diseñado para este fin. Como la palabra cuestionario ya se emplea para el módulo *quiz*, el término español elegido para *questionnaire* fue *encuesta*, lo cual no es terminológicamente preciso. Para evitar confusiones, en lo sucesivo usaremos el término *quiz* para referirnos al módulo *cuestionario* en Moodle, el término *cuestionario* para referirnos al instrumento de recogida de datos, y el término *encuesta* para referirnos al módulo correspondiente de Moodle (que sirve para crear cuestionarios de recogida de datos).

Hecha esta aclaración, la instalación del módulo *encuesta* y su uso no reviste ninguna dificultad. La diferencia con el módulo *quiz* es que éste requiere utilizar un sistema de puntuación, lo cual entorpecería su funcionamiento y además impone restricciones importantes, como el requisito

de dar respuestas acertadas como retroalimentación para las preguntas cortas, limitación que se podría superar mediante el uso de expresiones regulares que limiten el input del usuario, y que es factible mediante la instalación de una extensión de Moodle, pero esto complicaría enormemente la implementación del cuestionario.

Algo que sí hemos echado en falta es la disponibilidad de restricciones sobre el input del usuario mediante expresiones regulares, ya que esta extensión sólo es aplicable al tipo de pregunta respuesta corta dentro del módulo *quiz*. Curiosamente, Moodle implementa expresiones regulares, pero por algún motivo, sólo se puede emplear en el módulo *lección*. La única restricción aplicable en el módulo *encuesta* al tipo de pregunta respuesta corta, es el número máximo de caracteres, lo cual no es suficiente en muchos casos.

4.2. Test.

En Tono (2003) se hace alusión a un aspecto metodológico que muchos proyectos no han tenido en cuenta en la elaboración de córpora de aprendices: el nivel de competencia de la lengua objeto: «Selection based upon external criteria such as school year or age does not necessarily ensure that the subjects selected are comparable in terms of language proficiency.» (p. 801). En nuestro caso, este aspecto lo hemos resuelto mediante el Oxford Placement Test (OPT) (Allen, 2004), una prueba de nivel centrada en el conocimiento gramatical (Grammar Test) y la comprensión auditiva (Listening Test) del inglés que consta de un total de 200 ítems. Optamos por la versión publicada en el 2004 ya que está calibrada con el *Common European Framework of Reference for Languages* (Marco Común Europeo de Referencia para

las Lenguas) (Council of Europe, 2001), pruebas de nivel de inglés internacionalmente reconocidas (IELTS, TOEFL) y otras que están directamente vinculadas a los niveles de referencia del Marco Común Europeo y del Association of Language Testers in Europe (ALTE) (Cambridge ESOL Main Suite, CELS y BEC). La implementación del OPT en Moodle se hizo mediante el módulo *quiz*, ya que, a diferencia del cuestionario, el test requiere un sistema de puntuación que permita el cálculo automático del resultado.

Tanto para el *Grammar Test* como para el *Listening Test*, elegimos la opción múltiple como tipo de respuesta para cada ítem, a excepción de las secuencias 21-50 y 66-90 del *Grammar Test* que conforman textos completos y no se pueden responder de forma aislada como el resto; por tanto, estas dos secuencias se incluyeron como *respuesta anidada* (Cloze).

4.3. Ensayo.

Si bien es cierto que cada instrumento diseñado en la plataforma Moodle cumple una función esencial en la elaboración del corpus, el ensayo es lo que realmente nos permite obtener muestras de producción escrita de aprendices de inglés. Además, en una fase posterior, a cada texto se le realizará un etiquetado morfosintáctico, así como una anotación de errores para el desarrollo de futuros estudios basados en *Computer-aided Error Analysis* (Dagneux, Denness, & Granger, 1998), cuyo fin es la identificación de errores universales, común a todos los aprendices, y errores específicos que deriven de la lengua materna del aprendiz.

Para los ensayos también optamos por el módulo *quiz*, que en este caso consta de una sola pregunta de tipo ensayo. En un principio,

nos planteamos implementarlo mediante el módulo de tarea: texto en línea, sin embargo este módulo sólo ofrece la opción de trabajar con texto en formato HTML y el tipo de texto con el que se requiere para el corpus es texto sin formato (*raw text*).

A la hora de seleccionar los temas del ensayo a desarrollar por los aprendices, tuvimos muy en cuenta la heterogeneidad del nivel de inglés y de las áreas de especialidad del alumnado; por tanto, dimos a elegir entre ocho temas que no exigieran poseer un nivel avanzado del idioma ni del contenido. Finalmente, optamos por ocho de los veinte temas más típicos de ensayo propuestos en *Essay Writing for English Tests* (Duigu, 2002), un libro de texto que ofrece una serie de pautas a seguir para desarrollar la destreza escrita en inglés. He aquí un ejemplo de un tema de ensayo que seleccionamos: «Tobacco and alcohol are drugs that cause addiction and health problems. Should they be made illegal? Or should all drugs be legalised?» (Duigu, 2002, p. 65).

Finalmente, queremos hacer alusión a las cuentas de usuario que quedan habilitadas tras la aportación de datos del alumnado; como mencionamos en el apartado 3.3, el alumnado debe registrarse para poder acceder al curso; estas cuentas se mantendrán habilitadas con vistas a que estos mismos alumnos puedan seguir aportando producciones escritas en inglés, lo que nos permitiría desarrollar estudios de tipo longitudinal, que actualmente siguen escaseando en el campo de estudio de corpora de aprendices (Granger, 2004; Granger, 2009).

5. Exportación e integración de datos.

El proceso de exportación de datos es crucial para nuestros fines, y no carece de

complejidad. Moodle ofrece dos posibilidades para extraer los datos generados en la plataforma. La exportación como tal sólo es posible para los bancos de preguntas, mientras que el resto de contenidos ha de hacerse mediante el sistema de copia de seguridad. Esto significa que, en principio, la única manera de utilizar los datos exportados es importarlos en otra (o la misma) implementación Moodle. Sin embargo, nuestra intención es distinta, ya que deseamos disponer de los datos para analizarlos con otras herramientas.

En cuanto a los cuestionarios, el módulo de *encuesta* es, como hemos dicho, una extensión no incluida en los módulos de Moodle, por lo que tiene sus propias peculiaridades. Entre otras cosas, se puede visualizar los resultados del cuestionario ítem a ítem de forma gráfica (en gráficas de barras) y también ver los datos numéricos. No obstante, lo que necesitamos para llevar a cabo nuestro propio estudio estadístico es exportar los datos. En este caso, el módulo ofrece esta funcionalidad mediante la exportación de los resultados en formato CSV, que podemos tratar previamente con cualquier editor de texto, hoja de cálculo, o sistema gestor de bases de datos para seleccionar y ordenar las categorías de datos de la manera que más nos interese antes de importarlos al paquete estadístico.

La situación es distinta cuando hablamos de exportar datos provenientes de una actividad *quiz*, que es como hemos implementado los otros dos instrumentos. El único modo que Moodle facilita para exportar los datos de actividades de alumnos es mediante la copia de seguridad, que nos permite elegir exactamente qué queremos exportar. Sin embargo, lo que este método genera es un archivo con extensión “.mbz”

que, en principio, sólo nos serviría para recuperar datos en caso de necesidad o bien para importarlos en otro sitio Moodle. En realidad, este archivo no es más que un archivo comprimido zip que contiene una serie de carpetas y archivos en formato XML, el formato estándar de intercambio de datos. De esta manera, no resulta complicado extraer los contenidos que nos interesan mediante algún software. En nuestro caso, desarrollamos un pequeño programa en Python que lee directamente el archivo zip, lo descomprime, extrae los datos necesarios y los vuelca a un archivo de texto en formato CSV, listo para ser importado en cualquier paquete de software para su posterior análisis.

6. Resultados.

Como mencionamos ya en el resumen de este artículo, nuestro objetivo era ofrecer una descripción de la infraestructura y los protocolos desarrollados para el uso de Moodle, un gestor de aprendizaje, como herramienta de investigación en el campo de estudio de córpora de aprendices. Nuestro interés por buscar y adoptar nuevas técnicas de investigación surge a raíz de experiencias previas en las que métodos más tradicionales fueron empleados; por tanto, en esta última sección exponemos el resultado de adoptar una plataforma como Moodle como herramienta de recogida de datos que, indiscutiblemente, repercute de manera muy positiva sobre fases posteriores del estudio.

El método tradicional impreso para la recogida de datos, además de ralentizar el proceso del estudio, puede repercutir sobre el grado de fiabilidad y validez de los datos en sí ya que no podemos encontrar con multitud de casos de letra ilegible de los sujetos en cuestionarios y ensayos, así como respuestas

en blanco o respuestas múltiples en el test de nivel. Además, teniendo en cuenta que cada test contiene un total de 200 ítems, a corrección manual no asegura en absoluto la misma rigurosidad y precisión en el cálculo del resultado final.

Dos de los aspectos que más ralentizan el proceso de recogida de datos son la corrección manual de los test y la transcripción de los ensayos. A pesar de que el OPT incluye una plantilla de corrección para agilizar dicho proceso, la automatización y precisión que nos proporciona un ordenador no es comparable con ningún otro sistema que requiera trabajo manual. Incluso haciendo uso de la plantilla de corrección, corregir un sólo test puede llevar, como mínimo, 10 minutos. Esto sería posible con una muestra como la que actualmente poseemos, 119 test, pero nuestra intención es seguir aumentando esta cifra para aumentar la representatividad de la misma y, en ese caso, seguir el método tradicional de recogida de datos no es factible con los medios humanos de que disponemos.

Teniendo en cuenta que nuestro objetivo general consiste en la creación de un corpus de aprendices de inglés como lengua extranjera, resulta indispensable que exista un control y una planificación exhaustivos en el proceso de recogida de datos ya que esto repercutirá de manera decisiva sobre el diseño, y por tanto validez, del corpus (Tono, 2003).

Si realmente pretendemos que los córpora de aprendices sirvan de base empírica para la creación o confirmación de hipótesis desarrolladas en los campos de adquisición y enseñanza del inglés como lengua extranjera (Granger, 2009; Römer, 2008; Tono, 2003; García-González, 2011) primero habrá que asegurar que esas bases sean fiables y válidas para que las conclusiones extraídas también lo sean.

7. Discusión.

Con todo esto no queremos dar a entender que el trabajo aquí descrito carezca por completo de trabajo manual ya que, lógicamente, todos los instrumentos de investigación tuvieron que ser informatizados. Más aún, la infraestructura y protocolos desarrollados para el uso de Moodle como plataforma de apoyo a la creación de un corpus no suponen la única alternativa al método impreso tradicional (Granger, 2003b); lo que sí es cierto es que Moodle representa una alternativa mucho más intuitiva, que no requiere de un conocimiento técnico avanzado en cuanto a la creación de contenidos y exportación de datos. Por tanto, esto podría suponer un puente factible entre el campo de la enseñanza y el campo de la investigación. Una de las razones por las que el uso del corpus de aprendices en el ámbito de la enseñanza no se ha extendido lo suficiente es por la falta de relevancia o aplicación general que tiene en el aula. El uso de una plataforma como Moodle permitiría a los propios docentes desarrollar materiales didácticos basados en producciones lingüísticas de los alumnos y, por tanto, aportarle relevancia a dicha aplicación. Si bien es cierto que el despliegue de una implementación propia, como se ha hecho en nuestro caso, imponga disponer de determinados recursos y conocimientos técnicos, esperamos que nuestra experiencia sirva de referente para aquellos investigadores o docentes que deseen adaptar una plataforma como Moodle a sus propios intereses.

8. Fuentes de financiación.

La presente investigación forma parte del Proyecto de I+D+i FFI2011-25893,

denominado «Lingmotif: Análisis de Sentimiento en Textos: Aplicaciones Didácticas y Financieras» dirigido por el Dr. Antonio Moreno Ortiz y financiado por el Ministerio de Ciencia e Innovación, Gobierno de España.

9. Referencias bibliográficas.

- Allen, D. (2004). *Oxford Placement Test: Marking Kit with User's Guide and Diagnostic Key*, vol. 1, 1-63. Oxford: Oxford University Press.
- Atkins, S., Clear, J. & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1-16. <http://dx.doi.org/10.1093/lc/7.1.1>
- Baker, R.S. & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining (JEDM)*, 1, 3-17.
- Brown, H.D. (2000). *Principles of language learning and teaching*. New York: Longman.
- Cestero, A.M., Penados, I., Blanco, A., Camargo, L. & Granda, J.S. (septiembre, 2001). Corpus para el análisis de errores de aprendices de E/LE (CORANE). Actas del XII Congreso Internacional de ASELE. (pp. 527-534). Valencia, España.
- Cobb, J. & Steele, C. (2011). *Association learning management systems 2011: Special Blackboard edition*. Recuperado de <http://www.blackboard.com/>
- Council of Europe (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Recuperado de http://www.coe.int/t/dg4/linguistic/cadre_en.asp
- Correa, J. I. (2005). La integración de plataformas de e-learning en la docencia universitaria: Enseñanza, aprendizaje e investigación con Moodle en la formación

- inicial del profesorado. *RELATEC. Revista Latinoamericana de Tecnología Educativa*, 4(1), 37-48.
- Dagneaux, E., Denness, S. & Granger, S. (1998). Computer-aided error analysis. *System: An International Journal of Educational Technology and Applied Linguistics*, 26, 163-174. [http://dx.doi.org/10.1016/S0346-251X\(98\)00001-3](http://dx.doi.org/10.1016/S0346-251X(98)00001-3)
- Dobozy, E. & Reynolds, P. (2010). From LMS to VLE or from supermarkets to airports: Classifying elearning platforms using metaphors. *Proceedings of the 5th International LAMS Conference 201*, 92-103. Recuperado de http://lams2010sydney.lamsfoundation.org/pdfs/papers/06_Dobozy_Final.pdf
- Dougiamas, M. (november, 2003). Content Management System. *MoodleForums*. Recuperado de <http://moodle.org/mod/forum/discuss.php?d=3383>
- Duigu, G. (2002). *Essay Writing for English Tests*. Australia: Academic English Press.
- Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.
- Estévez, M.E. (2012). Análisis y beneficios de la incorporación de las TIC en el área de lengua castellana y literatura: un caso práctico. *Píxel-Bit. Revista de Medios y Educación*, 40, 21-34.
- García-González, J.L. (2011). Evaluación automatizada de narraciones en el área de Lengua Extranjera. *Píxel-Bit. Revista de Medios y Educación*, 39, 29-40.
- Granger, S. (2002). A Bird's-eye view of learner corpus research. En S. Granger, J. Hung & S. Petch-Tyson (Eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam, Philadelphia: John Benjamins.
- _____ (2003a). Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal*, 20 (3), 465-480.
- _____ (2003b). The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 37(3), 538-546. <http://dx.doi.org/10.2307/3588404>
- _____ (2004). Computer Learner Corpus Research: Current Status and Future Prospects. En U. Connor, & T.A. Upton (Eds.). *Applied Corpus Linguistics: A Multidimensional Perspective, Language and Computers* (vol. 52, pp. 123-145). Amsterdam and Atlanta: Rodopi.
- _____ (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. En J. Aijmer (Ed.). *Corpora and Language Teaching, Studies in Corpus Linguistics* (vol. 33, pp. 13-32). Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Lengo, N. (1995). What is an error?. *English Teaching Forum*, 33 (3), 20-24.
- Mallon, D. (2011). *Learning Management Systems 2011: The Definitive Buyer's Guide to the Global Market for Learning Management Solutions*. Recuperado de <http://www.bersin.com/Store/Details.aspx?docid=103313427>
- O'Keeffe, A. & McCarthy, M. (2010). *The Routledge Handbook of Corpus Linguistics*. New York: Routledge.
- Pérez Hernández, C. (2002). *Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicos basados en el conocimiento*. Málaga: University of Málaga. Recuperado de <http://elies.rediris.es/elies18/>
- Roca, M.L., Piñeiro, M.L. & Gigirey, I. (2012). The problem of false friends in learner language: Evidence from two learner corpora.

En L. Pérez Ruiz, I. Parrado Román & P. Tabarés Pérez (Eds.). *Estudios de Metodología de la Lengua Inglesa* (vol. 5, pp. 35-44). Valladolid: Centro Buendía, Universidad de Valladolid.

Römer, U. (2008). Corpora and Language Teaching. En A. Lüdeling & M. Kytö (Eds.). *Corpus Linguistics: An International Handbook* (pp. 112-130). Berlin: Walter de Gruyter.

Romero, C. & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), 601-618. <http://dx.doi.org/10.1109/TSMCC.2010.2053532>

_____ & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384. <http://dx.doi.org/10.1016/j.compedu.2007.05.016>

Ros Martínez de Lahidalga, I. (2008). Moodle, la plataforma para la enseñanza y organización escolar. *Ikastorratza, e-Revista de Didáctica 2*. Recuperado de http://www.ehu.es/ikastorratza/2_alea/moodle.pdf

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10 (3), 209-231.

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning*, 2(1). Recuperado de http://itdl.org/Journal/Jan_05/article01.htm

Sinclair, J. (1991). *Corpus, Concordance, Collocation: Describing English Language*. Oxford: Oxford University Press.

_____ (2004). *Trust the text*. London: Routledge.

Tono, Y. (march, 2003). Learner corpora: design, development and applications. Presentado en *Corpus Linguistics 2003*

Conference (CL 2003). Lancaster, U.K. Recuperado de <http://ucrel.lancs.ac.uk/cl2003/>

Torruella, J. & Llisterri, J. (1999). Diseño de corpus textuales y orales. En J.M. Bleuca, G. Clavería, C. Sánchez Lancis & J. Torruella (Eds.). *Filología e informática. Nuevas tecnologías en los estudios filológicos* (pp. 45-77). Barcelona: Editorial Milenio.

Vicente, P. & Reis, E. (2007). Methodological Issues in Online Surveys. En M. Baptista Nunes, P. Isaías & J. Barroso (Eds.). *Proceedings of the IADIS International Conference on WWW/Internet* (vol. 2, pp. 173-176).

Fecha de recepción: 09-07-2012

Fecha de evaluación: 08-10-2012

Fecha de aceptación: 27-12-2012